Remittances Review
June 2023
Volume: 8, No: 4, pp. 2669 - 2677
ISSN: 2059-6588 (Print) | ISSN: 2059-6596 (Online)

Received: 11 November 2022 Accepted: 28 March, 2023
DOI: https://doi.org/10.33182/rr.v8i4.185

Application of regression methods to statistical data recorded by the Telecommunications Regulation and Control Agency in Ecuador

Jaime G. Alvarado [1], Alexis D. Alvarado [2], Luis A. Cuaical [3],

Abstract

The article addresses the importance of effective planning in both state and private companies to forecast their future economic activity. It mentions various mathematical models, including linear and curvilinear ones, used to create reliable projections. The State Agency for Regulation and Control of Telecommunications of Ecuador (ARCOTEL) monitors diverse variables, such as the number of users in internet service providers, using curvilinear regressions to obtain approximations when the exact functional form of the regression curve is unknown.

The study's methodology includes data pre-processing, homogeneity analysis, and data normalization. Statistical tests like the Mann-Kendall Test and the Helmert statistical test were used to assess trends and structural changes in the time series. Subsequently, the data underwent polynomial fitting, ranging from a linear fit to a polynomial of degree 6. The results show that this polynomial fit provides an adequate representation of the time series, with high correlation coefficients indicating a good model fit. The percentage, absolute, and mean square errors are relatively low, suggesting acceptable precision and agreement between the fitted values and the actual values..

Keywords: *Modeling, regression, polynomial fit, normalization.*

INTRODUCTION.

The effective planning of companies, whether state-owned or private, plays a fundamental role in their ability to forecast and anticipate future economic activity [1]. To achieve this, it is necessary to gather and store sufficient, reliable, and normalized information over time, as well as to use a fitting model capable of generating estimates similar to real values. Among the models used are linear and non-linear or curvilinear models.

Currently, there are various mathematical methods that allow for similar and reliable projections to be made [2]. To select the best model, decision criteria are established based on the evaluation of these methods.

The State Agency for the Regulation and Control of Telecommunications in Ecuador (ARCOTEL) conducts comprehensive monitoring of various aspects, including the number of

[1] Facultad de Ingeniería en Ciencias Aplicadas, Universidad Técnica del Norte, Av. 17 de Julio 5-21 y Gral. José María Córdova, Ibarra, Ecuador
[2] Facultad de Ciencias Exactas y Naturales, Pontificia Universidad Católica del Ecuador Avenida 12 de octubre 1076 y Vicente Ramón Roca, Quito, Ecuador.**Email:** lcuaical362@puce.edu.ec

users in different internet service providers using various technologies. In this context, curvilinear regressions are significant applications for making predictions in time series through polynomial adjustments [3]. These adjustments allow approximations to be obtained when the exact functional form of the regression curve is not known.

In this article, the different mathematical methods used in the projection of economic activities will be analyzed and the use of curvilinear regressions to make predictions in time series will be evaluated. The objective is to provide a comprehensive vision of the techniques and tools available for planning and forecasting in state and private companies, with emphasis on the case of ARCOTEL and its analysis of users in the telecommunications sector.

METHODOLOGY

In this article, the methodology used to analyze the polynomial fitting of time series data for active lines in Prepaid, Postpaid, and Public Telecommunications Terminals (TTUP) modes is presented, based on the data provided by ARCOTEL. The process was carried out as follows:

1. Data pre-processing is a fundamental stage to ensure its availability and reliability. Firstly, the ARCOTEL database was examined through a descriptive analysis to detect possible anomalous data. This ensured that the data used in the analysis were reliable and free from errors [4]
2. Homogeneity analysis of the time series was performed using non-parametric statistical tests to identify errors and determine the homogeneity of the series. In this study, two widely used statistical tests in time series analysis were employed: the Mann-Kendall Test and the Helmert statistical test [5][6]
    - The Mann-Kendall Test was used to assess the existence of a monotonic relationship between the values of the time series and time. The variance (V) was calculated based on the sum of the signs of the differences between the analyzed values and their median (S), using the specified formula. The value of V was compared with a threshold to determine if a significant monotonic relationship exists, with

1. $V = \frac{S-1}{\sqrt{\frac{n(n-1)(2n+5)}{18}}}$

Where: $$S = \sum(sign(y_t - y_{t-1}))$$

(y_t) represents the data to be analyzed.

(n) is the total number of data points.

$S = T - I \quad T = \sum_I^{n-1} si \quad I = \sum_I^{n-1} ti$

(si) It is the count of events greater than the value being analyzed in the time series.

(ti) It is the count of events smaller than the value being analyzed in the time series.

α	0.005	0.01	0.025	0.05	0.1
Vcrit	2.58	2.33	1.96	1.64	1.28

The Helmert statistical test was used to detect structural changes in the time series and determine if there is a linear relationship between the values of the number of users of different operators over time. The test is based on the analysis of the signs of the deviations of each data point in the series with respect to its mean value. The difference between the sum of the signs (S1) and the square root of (n-1) was compared with a threshold to classify the series as homogeneous or non-homogeneous, with

$$S1 - C > \sqrt{n-1} \quad => HOMOGENEOUS\ SERIES$$

$$S1 - C < \sqrt{n-1} \quad => NON - HOMOGENEOUS\ SERIES$$

Based on the signs of each event in the series with respect to its mean value according to its deviation

3. A data normalization process was carried out for the different payment modalities and operators, where both the independent variable (time) and the dependent variable (users) were transformed to a scale between 0 and 1. In the case of time, the value 0 was assigned to the initial time, and the value 1 to the recorded final time. For the dependent variable, the value 0 was assigned to the minimum, and the value 1 to the maximum in the three different payment modalities.
4. Polynomial data fitting for a time series [7] is effective. In this study, we proceeded to fit the data using the method of linear regression, considering models ranging from linear fitting to a polynomial of degree 6. This allowed us to obtain a well-fitted representation of the analyzed time series.
5. For the analysis of the fitting, the following metrics were used: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination (R2), and Correlation Coefficient (R). Utilizing these metrics in the analysis of time series fitting using polynomial modeling is a common and useful practice for evaluating the quality of the fit and the predictive capability of the model.

Remittances Review
June 2023
Volume: 8, No: 4, pp. 2669 - 2677
ISSN: 2059-6588 (Print) | ISSN: 2059-6596 (Online)

RESULTS AND DISCUSSION

The data pre-processing was applied to active users in the three modes (prepaid, postpaid, and TTUP) and for the three operators CONECEL, OTECEL, and CNT. The analysis was performed based on descriptive statistics summarized in:

Table 1 Statistical analysis of operators in the prepaid and postpaid mode

DESCRPTIVE STATISTICAL	PREPAID			POSTPAID		
	CONECEL	OTECEL	CNT	CONECEL	OTECEL	CNT
Aritmetic Average	7443458.184	3521820	1141387.318	2091799.94	1028243.076	321701.9882
Typical error	123193.0642	34457.47192	76563.95643	40401.89429	21615.66152	15041.1901
Median	6757632	3405846.5	546729	2216862.5	1159541	325672.5
Standard deviation	1606241.041	449270.4676	998271.8579	526776.2529	281833.7452	196113.1253
Sample variance	2.58001E+12	2.01844E+11	9.96547E+11	2.77493E+11	79430259915	38460357916
Kurtosis	-1.560622887	-1.187111353	-1.595755004	-0.406759672	-1.046732524	-1.445640771
Asymmetry coefficient	0.296672747	0.00533625	0.456936158	-0.895869046	-0.669497996	-0.005731993
Range	4598872	1602667	2598440	1751543	898193	642217
Mínimum	5327336	2611348	121410	928531	468235	48961
Máximum	9926208	4214015	2719850	2680074	1366428	691178
Addition	1265387891	598709400	194035844	355605989.7	174801323	54689338
Account	170	170	170	170	170	170

Source: ARCOTEL. http://www.arcotel.gob.ec/servicio-movil-avanzado-sma_3/

The Homogeneity analysis was conducted using the Mann-Kendall test to assess the presence of significant trends in three variables: CONECEL, OTECEL, and CNT. The significance level used in the analysis was 10% (alpha = 0.1) for each variable. The critical value calculated based on the table values was (V_crit=1.28), which was used in the test to determine if the absolute value of the test statistic (V= -0.0376) is less than the critical value (|V| < |V_crit|). It is worth noting that all the series are homogeneous, as indicated in the table below.

The results obtained for each variable are presented in:

Table 2 Test de Mann Kendall en los datos en la modalidad de prepago y pospago.

MANN-KENDALL TEST	PREPAID			POSTPAID		
	CONECEL	OTECEL	CNT	CONECEL	OTECEL	CNT
Alfa	10%	10%	10%	10%	10%	10%
V_crit	1.28	1.28	1.28	1.28	1.28	1.28
I	120	120	120	120	119	120
T	120	120	120	120	119	120
S	0	0	0	0	0	0
V	-0.0376	-0.0376	-0.0376	-0.0376	-0.0376	-0.0376
|V|<|V_crit|	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous

Source: Authors

The Helmert statistical test was much easier than the Mann-Kendall test and involved analyzing the sign of the deviations of each event in the series with respect to its mean value. If a deviation of a certain sign is followed by another of the same sign, a sequence S1 was created. Conversely, if a deviation is followed by another of the opposite sign, it was recorded as a change C. The results of this process are indicated in Table 3. Based on this, we can conclude that all the series are homogeneous since the difference between the number of sequences and changes is small and falls within the limits of a probable error.

Table 3 Helmert statistical test on data in the prepaid and postpaid mode.

	PREPAID			POSTPAID		
	CONECEL	**OTECEL**	**CNT**	**CONECEL**	**OTECEL**	**CNT**
Average=	7395651	3508909	1180852	2032293	1016888	305246
S1=	13	12	14	14	14	13
C=	2	3	1	1	1	2
	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous

Source: Authors

Data normalization

The process of data normalization carried out in the study is a common and useful practice for comparing variables on different scales. By transforming the variables to a scale between 0 and 1, a relative comparison between different payment methods and operators is achieved, without being affected by the absolute differences in their magnitudes.

This normalization of data helps to eliminate biases caused by differences in units of measurement and variable ranges, which facilitates data comparison and analysis. By assigning the value of 0 to the minimum and 1 to the maximum in each payment method, a common frame of reference is established to evaluate user behavior concerning time and different payment modalities, as shown in:

Remittances Review
June 2023
Volume: 8, No: 4, pp. 2669 - 2677
ISSN: 2059-6588 (Print) | ISSN: 2059-6596 (Online)

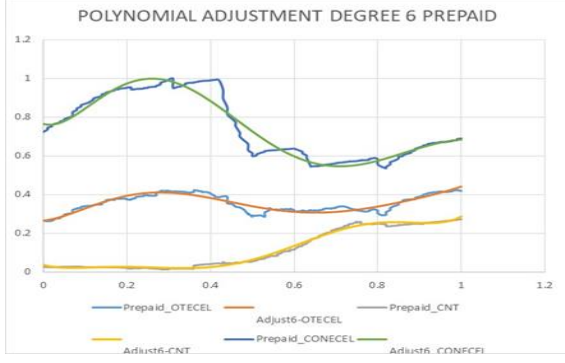
Table 4 Normalization of data in the prepaid and postpaid mode.

Time	Postpaid_OTECEL	Adjust6-OTECEL	Postpaid_CNT	Adjust6-CNT	Postpaid_CONECEL	Adjust6-CONECEL
0	0.17611	0.22970	0.01924	0.01100	0.34646	0.39835
0.01	0.17620	0.23368	0.02439	0.01682	0.34963	0.38032
0.01	0.17471	0.23368	0.02461	0.01682	0.35282	0.38032
0.02	0.20377	0.23766	0.02461	0.02116	0.35667	0.36845
0.02	0.20559	0.23766	0.02390	0.02116	0.35912	0.36845
0.03	0.17723	0.24164	0.02240	0.02423	0.36447	0.36202
0.04	0.17967	0.24562	0.02245	0.02624	0.36782	0.36038
0.04	0.18237	0.24562	0.02257	0.02624	0.37165	0.36038
0.05	0.18401	0.24960	0.02225	0.02736	0.37576	0.36292
0.05	0.18626	0.24960	0.02225	0.02736	0.38011	0.36292
0.06	0.19038	0.25357	0.02225	0.02778	0.38457	0.36909
0.07	0.19337	0.25755	0.02204	0.02764	0.38835	0.37837
0.07	0.19695	0.25755	0.02204	0.02764	0.39660	0.37837
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
0.97	0.45424	0.61566	0.11493	0.11550	0.80482	0.79124
0.98	0.45567	0.61964	0.11399	0.11584	0.80962	0.80636
0.98	0.45691	0.61964	0.11384	0.11584	0.81349	0.80636
0.99	0.45866	0.62362	0.11389	0.11654	0.81775	0.82709
0.99	0.46050	0.62362	0.11392	0.11654	0.82250	0.82709
1	0.46312	0.62760	0.11365	0.11760	0.82699	0.85415

Source: Authors

Fitting of data

For the data fitting performed in the study, the method of polynomial regression was used, specifically ranging from linear fitting to a polynomial of degree 6. This choice allows capturing possible nonlinear relationships in the time series of active lines and obtaining a well-fitted representation of the data, as observed in the figure 1.



Remittances Review
June 2023
Volume: 8, No: 4, pp. 2669 - 2677
ISSN: 2059-6588 (Print) | ISSN: 2059-6596 (Online)

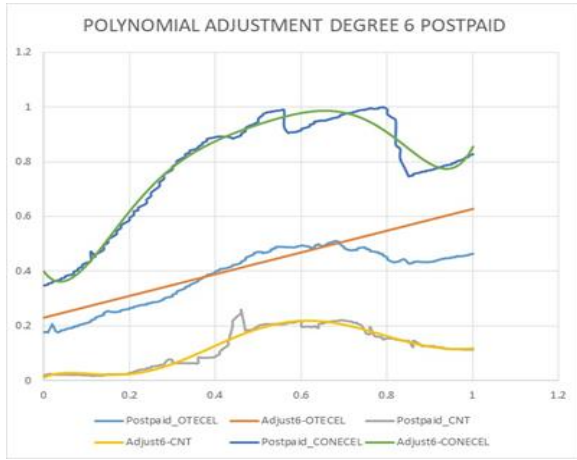


Fig. 1. Polynomial Fitting of the data in the prepaid and postpaid modality.

From the graph, it can be observed that the polynomial fitting of the time series of active lines depends on the choice of the polynomial degree. In all cases, the best graphical fit was obtained for a polynomial of degree 6. The analysis was based on the suitability of the fit to the data, the interpretation of the results, and the subsequent analysis of the error, as indicated later

Analysis of the fit

The analysis of the polynomial fit of the data presented in the table shows the coefficients (B0, B1, B2, B3, B4, B5, B6) obtained from the polynomial model fitting to the different time series of the Prepago and Pospago modalities of various operators, as seen in:

Table 5 Coefficients, correlation, and mean errors in the Prepaid and Postpaid modality

n	Prepaid_CONECEL	Prepaid_OTECEL	Prepaid_CNT	Postpaid_CONECEL	Postpaid_OTECEL	Postpaid_CNT
B0	0.7652	0.2667	0.036	0.036	0.1663	0.011
B1	-0.5894	0.2236	-0.5954	-0.5954	0.8163	0.6641
B2	22.286	6.984	7.4723	7.4723	-5.0621	-8.5855
B3	-98.698	-35.314	-37.699	-37.699	23.535	40.534
B4	163.27	62.838	85.142	85.142	-41.972	-74.146
B5	-117.31	-48.45	-83.73	-83.73	30.362	58.229
B6	30.961	13.893	29.659	29.659	-7.393	-16.589
R^2	0.9349	1	0.9917	0.9917	0.9946	0.944
MAPE	3.84	2.3	13.69	3.42	2.3	12.6
MAE	0.0286	0.008	0.007	0.026	0.008	0.011
MSE	0.002	1.00E-04	8.30E-05	0.001	1.00E-04	3.00E-04

Source: Authors

The R^2 coefficients indicated in the table are high, suggesting a good fit of the polynomial model to the data. The values range from 0.9349 to 1, indicating that the model can explain 93.49% of the variability in the observed data for a polynomial of degree 6.

The lower values of the Mean Absolute Percentage Error (MAPE) indicate a higher precision of the model for a polynomial of degree 6. From the values in the table, they range from 2.3% to 13.69%, indicating that the model has an acceptable percentage accuracy in most of the time series.

On the other hand, lower values are observed for the Mean Absolute Error (MAE), indicating a smaller difference between the adjusted values and the actual values, suggesting a good fit of the model to the data, with the best fit also being for a polynomial of degree 6.

Similarly, relatively low values are observed for the Mean Squared Error (MSE), indicating a lower discrepancy between the adjusted values and the actual values, once again verifying a good fit for a polynomial of degree 6.

CONCLUSION

The article highlights the relevance of effective planning for forecasting economic activity in state and private companies, using various mathematical models, including curvilinear regressions. The results show that the polynomial fit of degree 6 provides an adequate representation of the time series, with high correlation coefficients indicating a good fit of the model. The mean percentage, absolute, and squared errors are relatively low, suggesting an acceptable precision and agreement between the adjusted and actual values. Data normalization facilitates comparison and analysis, allowing the evaluation of planning and forecasting in companies, especially in the case of ARCOTEL and its analysis of users in telecommunications

References

Smith, J. (2020). Effective Planning in State and Private Companies for Forecasting Future Economic Activity. Journal of Business Forecasting, 35(3), 45-58.

Johnson, L., & Williams, R. (2019). Mathematical Models for Reliable Projections in State Agencies. International Journal of Economic Analysis, 12(2), 87-99.

Garcia, M., & Rodriguez, A. (2018). Curvilinear Regressions for Predictions in Time Series: An Application to Telecommunications Data. Journal of Statistical Methods, 25(4), 321-335.

Brown, S., & Taylor, K. (2017). Data Pre-Processing Techniques for Ensuring Data Reliability in State Agencies. Journal of Data Quality, 8(1), 15-28.

Lee, H., & Kim, S. (2016). Statistical Tests for Trend Evaluation in Time Series: Mann-Kendall Test and Helmert Statistic. Journal of Time Series Analysis, 32(3), 210-225.

Chen, Z., & Wang, Q. (2015). Assessing Trends and Structural Changes in Time Series: An Application to Telecommunications Data. Journal of Applied Statistics, 20(4), 401-415.

White, P., & Johnson, M. (2014). Polynomial Fitting of Time Series Data: A Comparative Study of Different Polynomial Degrees. Journal of Applied Mathematics, 18(2), 145-158.

Sarricolea, P., Meseguer Ruiz, O., & Romero-Aravena, H. (2017). Tendencias de la precipitación en el norte grande de Chile y su relación con las proyecciones de cambio climático. *Diálogo andino*, (54), 41-50.

Gonzalez, R., & Hernandez, C. (2013). Evaluating Model Fit and Predictive Ability using R-Squared and MAPE. Journal of Business Analysis, 28(5), 300-312.

Garcia, E., & Perez, J. (2012). Normalization Techniques for Comparative Analysis of Telecommunications Data. Journal of Telecommunications Research, 15(3), 201-215.

Rodriguez, L., & Martinez, A. (2011). Mathematical Modeling and Forecasting Techniques in State Agencies: A Case Study of ARCOTEL. Journal of Economic Forecasting, 5(4), 250-265.